



Cold Start on Online Advertising Platforms: Data-Driven Algorithms and Field Experiments

Renyu (Philip) Zhang

(Joint work with Zikun Ye, Dennis J. Zhang, Heng Zhang, Xin Chen)

Online Advertising Platform

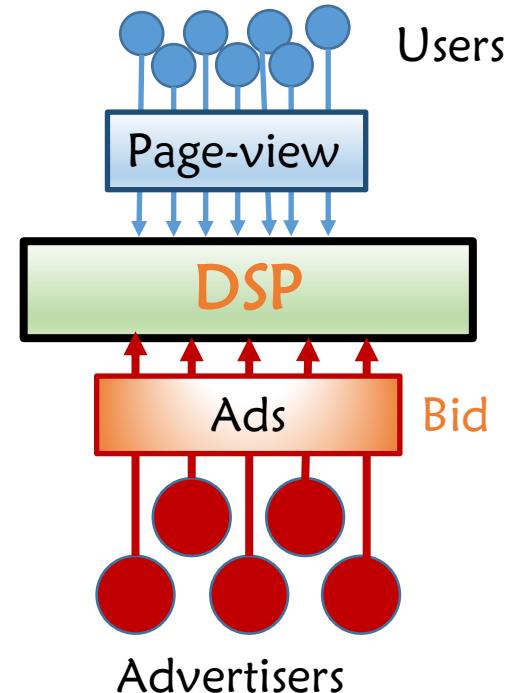
- Online advertising platform: Demand Side Platform (DSP)

- Fundamental question of a DSP:

When an ad request (page view) arrives,
which ad should be displayed to her?

- Core business logic:

The DSP runs large-scale **auctions** to
determine which ad to display, in order to
maximize the **platform's revenue**.



Performance-Based In-Feed Ads



Ad Impression on the Platform



External Site of Advertiser

DSP and Advertisers

- Performance-based ads: More **conversions/actions** at a low cost (i.e., high **Return on Investment**)
 - Mobile games: Activation & deposit
 - eCommerce: Activation & purchase
- Auction Mechanism & Billing Option: Optimized Cost-Per-Click (**oCPC**)
 - The advertisers **bid on conversions** and **pay upon clicks** (a compromise between the advertisers and the DSP).
 - The ads are ranked by the **expected cost-per-mile (eCPM)**, which is the expected **revenue per unit impression**.
 - The impression is allocated to the ad with the **highest eCPM**.
 - First-price auction.

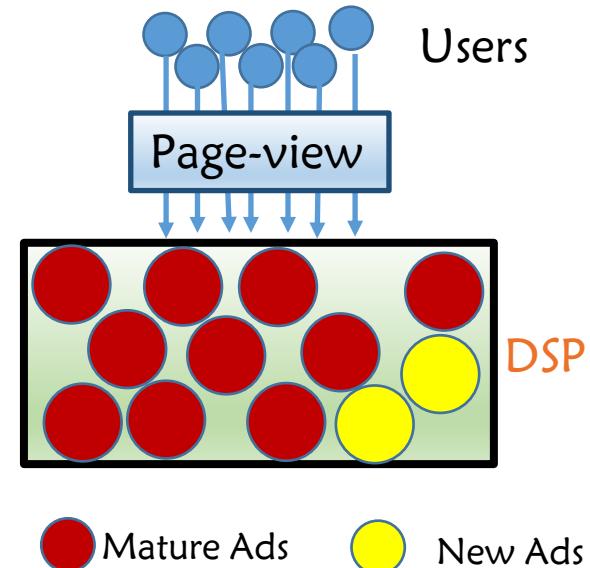
Mechanism	Charged-upon	Fee-Deduction	Rank-by
oCPC	click	$pCVR * bid_convert$	$eCPM = pCTR * pCVR * bid_convert$

pCTR (predicted CTR) and pCVR (predicted CVR conditioned on click) are produced by underlying deep neural networks of the DSP

- The bids are more like target costs-per-action (CPA), instead of real costs.
 - The platform adjusts the bid for conversions in real-time through the Proportional-Integral-Derivative (PID) controller to control the realized CPA.

Cold Start on DSP

- **New ads:** No sufficient data to estimate pCTR and pCVR accurately.
 - Unclear revenue implications for the DSP.
 - Successful cold start of new ads **thickens the ad pool** and **boosts advertiser retention**.
- **Mature ads:** High and stable revenues, minimally affecting user experiences.



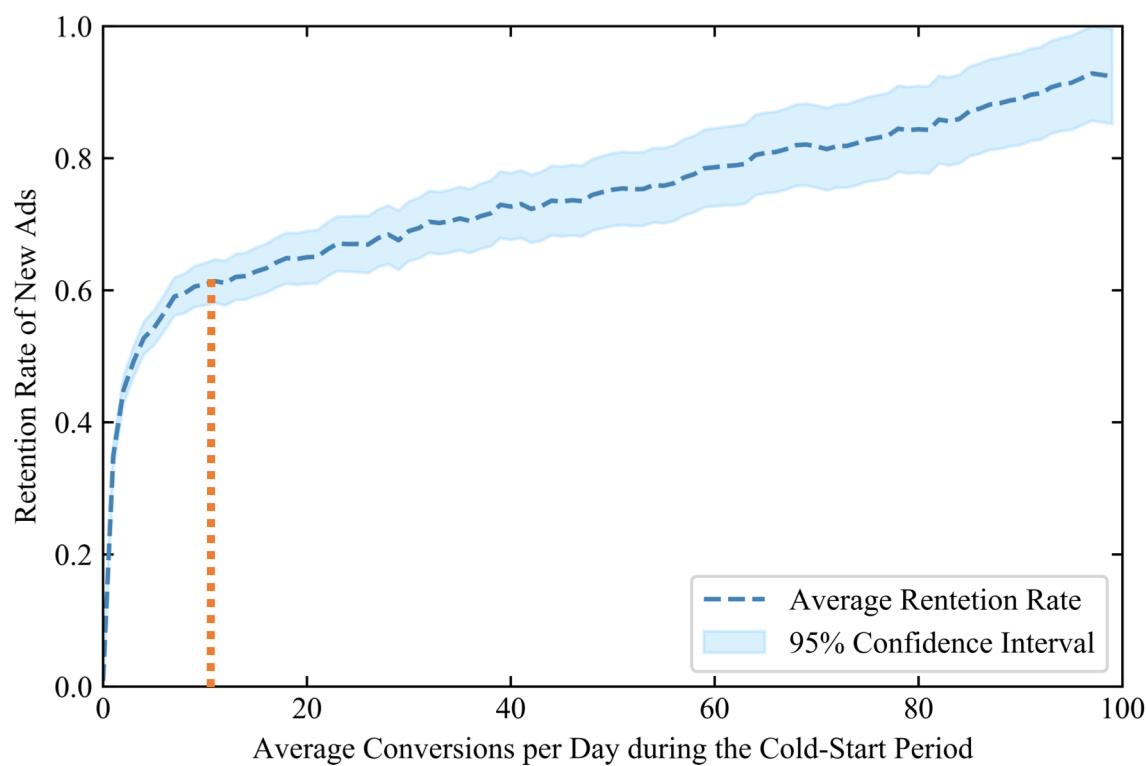
Core Problem: How to allocate user impressions between new and mature ads to balance the **short-term revenue** and the **long-term cold start value?**

Exploitation

vs

Exploration

Cold Start on DSP: Ad Retention



- X-axis: # of conversions in the first 3 days.
- Y-axis: Proportion of ads that will stay active on the DSP every single day in the next 2 weeks.
- Every thing is rescaled.

- **Key observation:** If the # of conversions in the cold start period surpasses 10, the long-term retention and value soon flatten.
- The # of new ads with # of conversions ≥ 10 is a key metric to evaluate ad cold start effectiveness.
- The phenomenon of the left figure is robust with respect to different definitions of ad retention.

Key Research (and Business) Questions

- With inaccurate predictions of CTR and CVR for new ads, how to smartly balance short-term revenue and long-term market thickness (i.e., cold start value)?

Primal-dual based MAB algorithms:
Shadow Bidding with Learning (SBL)

- With different ads competing for the same page-views, how to unbiasedly estimate the value of our proposed algorithm?

User-ad two-sided experiment framework



Related Literature

- **Ad Cold Start:** More accurate CTR and CVR predictions for new ads.
 - Dave and Varma (2012, 2014), Zhou et al. (2018), Choi et al. (2020), etc.
- **Contextual Bandits:** Establishing sublinear regret bounds with **optimization oracle**.
 - Langford and Zhang (2007), Chu et al. (2011), Bandanidiyuru et al. (2013), Agrawal et al. (2014), Agrawal et al. (2016), Jacot et al. (2018), Arora et al. (2019), etc.
- **Operations problem with online learning:** Bandit learning algorithms applied to **revenue management and inventory** problems.
 - Besbes and Zeevi (2009), Nambiar et al. (2019), Chen et al. (2019), Golrezaei et al. (2019), etc.
- **Experimental evaluations of algorithms on two-sided platforms:** Debiasing the estimate when **SUTVA does not hold**.
 - Ha-Thuc et al. (2020), Johari et al. (2020), etc.

Highlight of Main Contributions

- Modeling:

- Formulation of ad cold start as a **data-driven optimization model** under uncertainty

- Algorithm:

- SBL algorithm: **Duality + MAB + neural networks**
 - Bridging the gap between **learning theory** and **online advertising practice**

- Experiment:

- **Two-sided experiment** to evaluate a bandit algorithm
 - Novel experiment design to **restore SUTVA**
 - **Unbiasedly demonstrate the significant value** of the SBL algorithm in ad cold start



Model for Cold Start

- Problem setting: First-price auction, $\text{CVR} = \text{pCVR} = 1$ ($\text{oCPC} = \text{CPC}$)
- A set of new ads $A := \{1, 2, \dots, K\}$, with bids $\{b_1, b_2, \dots, b_K\}$
- A set of user impressions, arriving at the DSP sequentially: $[T] := \{1, 2, \dots, T\}$
- Context associated with each user: $x_t \in X$, i.i.d. on a countable set X , $x_t \sim \mathcal{D}$
- $v_{ij} = 0, 1$: Whether a user with context_i clicks ad_j; $y_{tj} = 0, 1$: Whether display ad_j to user_t
- CTR: $c_{ij} := E[v_{ij}]$, \hat{c}_{ij}^t := pCTR estimate by the underlying DNN in round t
- Sequence of events in each round t:



Reward Upper Bound and Regret

- Our objective is to identify a policy π to maximize the expected reward $E[\Gamma]$:

$$\Gamma := \underbrace{\sum_{j=1}^K b_j \left(\sum_{t=1}^T v_{x_t,j} y_{tj} \right)}_{\text{Short-term revenue}} + \underbrace{\sum_{j=1}^K \beta_j \min \left\{ \sum_{t=1}^T v_{x_t,j} y_{tj}, \alpha T \right\}}_{\text{Long-term cold start value}}$$

- β_j is the cold start value of ad j , determined by **business sense** and **simulation**.
- αT is the threshold for cold start success.

Lemma (Fluid upper bound). We have the following upper bound for the expected reward:

$$\frac{1}{T} E_{\mathcal{D}, \pi} [\Gamma] \leq \text{OPT} := \max_{y_i \in \Delta_A} \left\{ \sum_{j=1}^K E_{i \sim \mathcal{D}} [c_{ij} y_{ij} b_j] + \sum_{j=1}^K \beta_j \min \left\{ E_{i \sim \mathcal{D}} [c_{ij} y_{ij}], \alpha \right\} \right\}$$

where Δ_A is the distribution over all the ads and $y_i = (y_{i1}, y_{i2}, \dots, y_{iK}) \in \Delta_A$ is the ad assignment distribution for a user with context i .

- The regret of a policy: $\text{Reg}(\pi) := T \cdot \text{OPT} - E_{\mathcal{D}, \pi} [\Gamma]$
- We seek to design an algorithm with **sublinear regret** and **implementable on a real DSP**.

Empirical Reward and Ad Allocation

- Empirically optimal ad allocation policy at round t :

$$\begin{aligned}
 & \max_{y_{sj} \geq 0, \sum_{j \in A} y_{sj} \leq 1} \underbrace{\sum_{s \leq t-1} \sum_{j \in A} \hat{c}_{ij}^s b_j y_{sj}}_{\text{Short-term revenue}} + \sum_{j \in A} \beta_j \min \left\{ \sum_{s \leq t-1} \hat{c}_{ij}^s y_{sj}, \alpha(t-1) \right\} \\
 & \quad \text{eCPM of ad } j \text{ in round } s
 \end{aligned}$$

Long-term cold start value

- Linearize:

$$\begin{aligned}
 & \max_{y_{sj} \geq 0, u_j \geq 0} \sum_{s \leq t-1} \sum_{j \in A} \hat{c}_{ij}^s b_j y_{sj} + \sum_{j \in A} \beta_j [\alpha(t-1) - u_j] \\
 & \text{s.t. } \sum_{j \in A} y_{sj} \leq 1, \forall s \leq t-1, \quad \sum_{s \leq t-1} \hat{c}_{ij}^s y_{sj} + u_j \geq \alpha(t-1), \forall j \in A
 \end{aligned}$$

- u_j is the number of conversions below the threshold.
- The model has a too high dimension to solve efficiently online.

Duality and Shadow Bidding

$$\min_{\lambda_j \in [0, \beta_j]} \sum_{s \leq t-1} \max_{j \in A} \left\{ \underbrace{\hat{c}_{ij}^s(b_j + \lambda_j)}_{\text{Adjusted eCPM for ad_j}} \right\} - \alpha(t-1) \sum_{j \in A} \lambda_j$$

- Only K decision variables: Can be efficiently solved using sub-gradient descent!
- λ_j is the dual variable for the cold start reward constraint: $\sum_{s \leq t-1} \hat{c}_{ij}^s y_{sj} + u_j \geq \alpha(t-1)$
 - We call λ_j the **shadow bid** of ad_j.
- The shadow bid λ_j is bounded by the cold start value of ad_j, β_j
- Actionable Insight:

- Smartly computing the **shadow bid** of each new ad and incorporating it into the auctions of the DSP could effectively **trade off short-term revenues with long-term cold start values!**
- Naturally fit into the ad auction system of a DSP in **practice**.

Shadow Bidding with Learning (SBL)

Shadow Bidding with Learning (SBL) Algorithm

- Update shadow bids at rounds τ_1, τ_2, \dots , where $\tau_{m+1} - \tau_m = \tau_m - \tau_{m-1} = O(T^{\frac{2}{3}})$
- For each round $t=1, 2, 3, \dots, T$
 1. Observe the context $x_t = i$. With probability $\epsilon_t = t^{-\frac{1}{3}}(K \log t)^{\frac{1}{3}}$, explore uniformly at random; with probability $1 - \epsilon_t$, display the ad $\operatorname{argmax}_j \hat{c}_{ij}^t(b_j + \lambda_j)$, with an arbitrary tie-breaking rule.
 2. If $t = \tau_m$, solve the empirical dual program to update λ , and update $m \leftarrow m + 1$
 3. Observe the click-through outcome, and update \hat{c}_{ij}^{t+1}

Existing approaches in the literature:

- Based on **empirical risk minimization oracle**.
- Regret benchmarked with **the best policy in a policy set**, dependent on the **policy set size**.
- Of **theoretical** nature, **not scalable** and **implementable** on a real DSP.

vs

SBL Algorithm:

- **Dual + MAB (epsilon greedy) + ML Oracle**.
- Optimal primal ad allocation with the **dual solution**, which can be obtained **efficiently**.
- Leveraging the underlying **ML Oracle**.
- **Implementable** on a large-scale DSP in practice with **minimal changes** to the system.

Theoretical Performance Guarantee of SBL

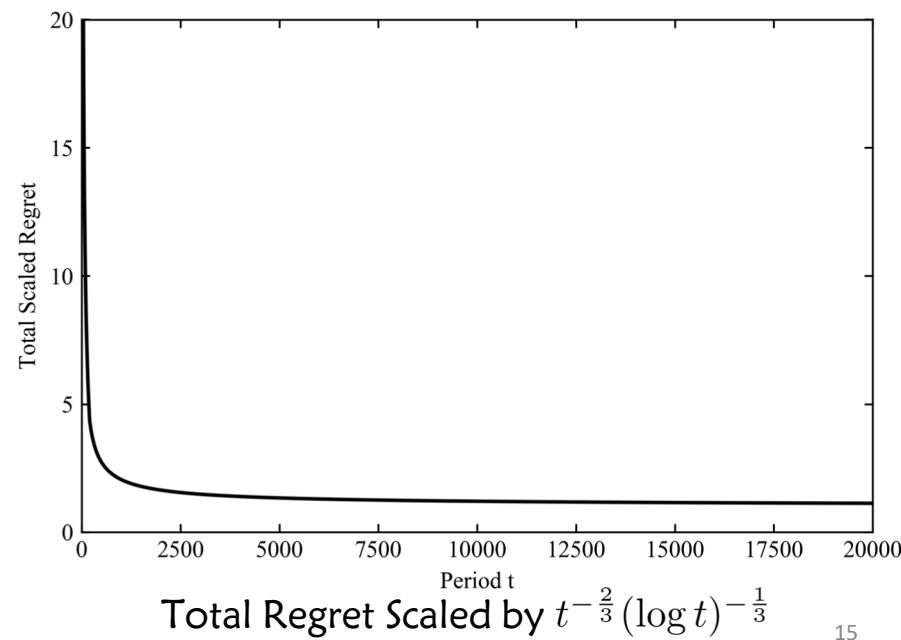
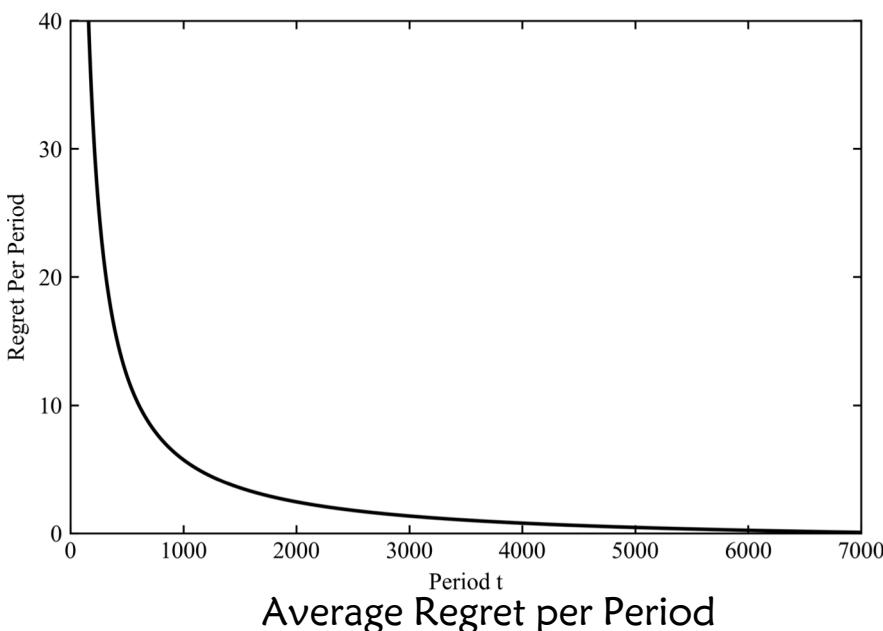
- **Machine learning oracle assumption:** With probability at least $1 - \delta$, the estimate \hat{c}_{ij}^t satisfies

$$|\hat{c}_{ij}^t - c_{ij}| \leq O\left(\sqrt{\log(1/\delta)d/n_j^t}\right)$$

where d is the dimension of the underlying machine learning oracle to obtain the pCTR \hat{c}_{ij}^t , n_j^t is the number of impressions for ad j by round t .

- Satisfied by (i) linear regressions; (ii) regression trees; and (iii) fully connected neural networks.

Theorem (Regret bound). The expected regret of SBL is bounded by $O(T^{\frac{2}{3}}K^{\frac{1}{3}}(\log T)^{\frac{1}{3}}d^{\frac{1}{2}})$.



Sketched Regret Analysis

- Key challenges:
 - History-dependent cold start reward (i.e., the knapsack bandit setting)
 - Dual-based bidding strategy implemented on the primal space
 - Regret dependent on the underlying machine learning oracle to predict CTR
 - Too high variance with Inverse Propensity Score to estimate the expected reward
- Key ideas and the road map to overcome the challenges:
 - Establish approximate complementary slackness and bound the duality gap between the empirical primal and the empirical dual, due to tie breaking in SBL by $O(T^{\frac{1}{3}}(\log T)^{\frac{1}{3}}K^{\frac{1}{3}})$
 - Build an auxiliary reward process independent of history: Each click of ad j generates a reward of $b_j + \beta_j$, irrespective of whether the threshold αT is met. Under SBL, bound the gap between the auxiliary reward process and the optimal reward by $O(T^{\frac{2}{3}}(\log T)^{\frac{1}{3}}K^{\frac{1}{3}}d^{\frac{1}{2}})$
 - Under SBL, bound the gap between the auxiliary reward process and the true reward process by $O(T^{\frac{1}{2}}(\log T)^{\frac{1}{2}}K^{\frac{1}{2}})$
 - Putting the above bounds together yields the expected regret of SBL is $O(T^{\frac{2}{3}}(\log T)^{\frac{1}{3}}K^{\frac{1}{3}}d^{\frac{1}{2}})$

Online Implementation of SBL (oSBL)

Online Shadow Bidding with Learning (SBL) Algorithm

- Update shadow bids at rounds τ_1, τ_2, \dots , where $\tau_{m+1} - \tau_m = 1$ hour, set $\alpha T = 10$, $\beta_j = 2b_j$
- For each round $t=1, 2, 3, \dots, T$
 1. Observe the context $x_t = i$. Choose top 150 (new & mature) ads and another 15 randomly picked new ads to join the auction.
 2. Obtain $\hat{c}_{ij}^t = pCTR * pCVR$. Display $\arg\max_{j \in [K_t]} \hat{c}_{ij}^t (b_{tj} + \lambda_j)$, where b_{tj} is the system bidding price calculated by the real-time PID system and $[K_t]$ is the set of the 165 ads who join the auction.
 3. If $t = \tau_m$, sample 4% of the auctions in the past hour \mathcal{H}_t and update the shadow bids λ by

$$\min_{\lambda_j \in [0, \beta_j], \forall j \in [K], \lambda_j = 0, \forall j \in [K']} \sum_{s \in [\mathcal{H}_t]} \max_{j \in [K_s]} \{\hat{c}_{ij}^s (b_{sj} + \lambda_j)\} - \alpha |\mathcal{H}_t| \sum_{j \in [K_s]} \lambda_j$$
 where $[K]$ is the set of new ads and $[K']$ is the set of mature ads
 4. Observe the click-through outcome, and update \hat{c}_{ij}^{t+1}

- Conversions are incorporated into the algorithm (CVR and pCVR are much smaller than 1).
- For a **mature ad**, the shadow bid is **0**.
- Shadow bids are updated **every hour** based on **sampled data**, implemented on top of the **PID** system.
- More like **uniform exploration** than epsilon-greedy.
- Cold start value is set at twice as much as the target CPA of the ad: $\beta_j = 2b_j$

Implementing and Testing SBL

- SBL algorithm implemented on a large-scale video-sharing platform (Platform O)
- How to **unbiasedly** evaluate the SBL algorithm?
- Naive **one-sided experiment** designs:
 - Ad-side randomization:

	Treatment New Ads	Control New Ads	Non-Experiment New Ads	Mature Ads
100% UV	Treatment Condition	Control Condition		

- User-side randomization:

		100% New Ads	Mature Ads
Treatment UV	Treatment Condition		
Control UV	Control Condition		
Non-Experiment UV			

Treatment = oSBL algorithm; Control = baseline algorithm (uniformly increase the bidding price of all new ads)

Violation of SUTVA

- Ad-side randomization:

	Treatment New Ads	Control New Ads	Non-Experiment New Ads	Mature Ads
100% UV	Treatment Condition	Control Condition		

- User-side randomization:

	100% New Ads	Mature Ads
Treatment UV	Treatment Condition	
Control UV	Control Condition	
Non-Experiment UV		

- SUTVA** (Stable unit treatment valuation assumption): The assignment of one unit to treatment or control will not affect the outcome of another unit.
- Ad-side randomization: 120% overestimate**
 - Violation of SUTVA: New ads from different groups compete on the same user impressions.
- User-side randomization: 40% underestimate**
 - Violation of SUTVA: The effect of oSBL spills over to the control group users.

Two-Sided Experiment Design

- A novel two-sided experiment design:

	20% Treatment New Ads	20% Control New Ads	60% Non-Experiment New Ads	Mature Ads
33% Treatment UV	B11	B12	B13	B14
33% Control UV	B21	B22	B23	B24
33% Non-Experiment UV	B31	B32	B33	B34

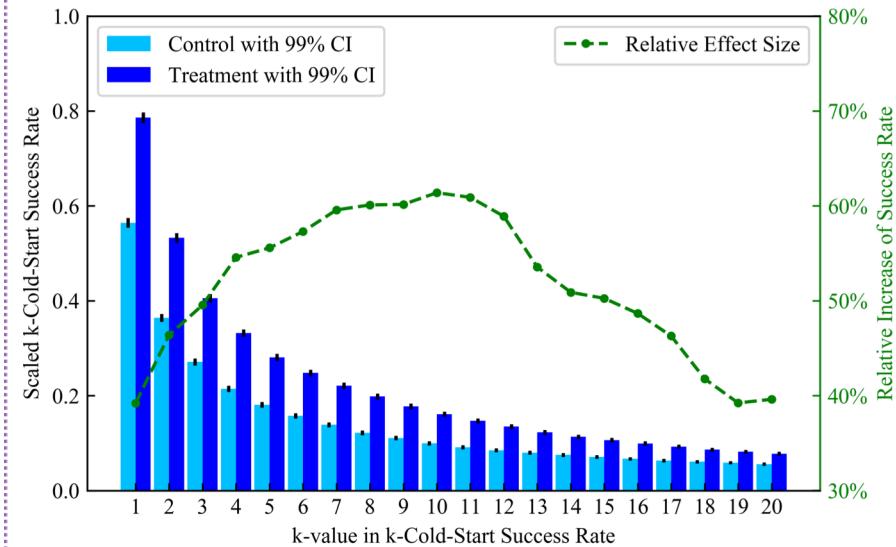
- Blue = oSBL, white = baseline algorithm, grey = ad blocked (to remove externalities)
- SUTVA restored! Able to generate unbiased estimates.
- Impact on long-term cold start value: Comparing B11 with B22
- Impact on short-term revenue: Comparing B11+B13+B14 with B22+B23+B24
- Experiments conducted between May 23, 2020 and May 30, 2020

Experiment Results

- oSBL algorithm has caused:

Metric of Interest	Cold start success rate	Cold start reward	Total objective value	Total short-term revenue	CTR Prediction AUC
Relative Change	+61.62%***	+47.71%***	+0.147%***	-0.717%**	+7.48%*

- Given the gigantic scale of Platform O, the experiment implies a sizable **hundred-million-dollar boost** in ad revenue per year.
- Two-sided network effects** would give rise to a **much stronger increase** in the total long-term value of oSBL.
- Results are **robust** if we run regression models and control user- and ad- specific features.



Robustness of Results

Takeaways

- **SBL Algorithm:** A smart algorithm to bridge the gap between bandit learning theory and the ad cold start practice.
- **Two-sided experiment:** Unbiasedly estimate the value of SBL for a large-scale online advertising platform.
- SBL and two-sided experiment have the potential to optimize and evaluate more general **recommender systems of online two-sided platforms.**



Thank You!

Questions?